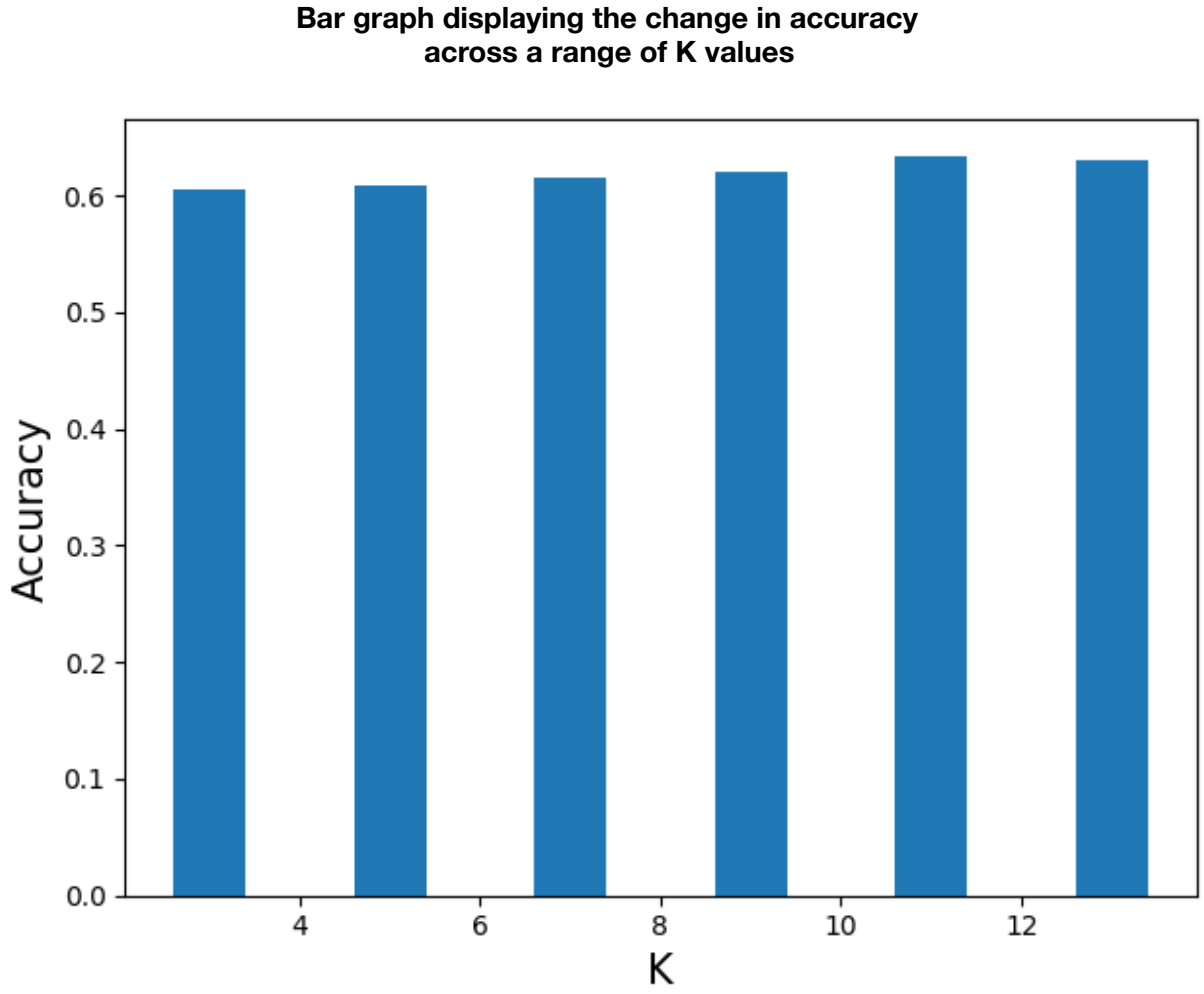
Q1 - KNN and Model Selection



Q2 - KNN and Model Selection

Kernel: linear C: 0.01 Accuracy: 0.8437005202442887
Kernel: linear C: 1 Accuracy: 0.8548970821081203
Kernel: linear C: 100 Accuracy: 0.854670888939154

Kernel: poly C: 0.01 Gamma: 0.1 Degree: 2 Accuracy: 0.8301289301063108
Kernel: poly C: 0.01 Gamma: 0.1 Degree: 3 Accuracy: 0.82650983940285
Kernel: poly C: 0.01 Gamma: 0.1 Degree: 5 Accuracy: 0.8172359194752319
Kernel: poly C: 0.01 Gamma: 1 Degree: 2 Accuracy: 0.8451707758425695
Kernel: poly C: 0.01 Gamma: 1 Degree: 3 Accuracy: 0.8515041845736259

Kernel: poly C: 0.01 Gamma: 1 Degree: 5 Accuracy: 0.8317122822890749
Kernel: poly C: 0.01 Gamma: 10 Degree: 2 Accuracy: 0.8504863153132776

Kernel: rbf C: 0.01 Gamma: 0.1 Accuracy: 0.8282062881700972
Kernel: rbf C: 0.01 Gamma: 1 Accuracy: 0.7894141596923773
Kernel: rbf C: 0.01 Gamma: 10 Accuracy: 0.7717710925130061
Kernel: rbf C: 1 Gamma: 0.1 Accuracy: 0.8437005202442887
Kernel: rbf C: 1 Gamma: 1 Accuracy: 0.8338611173942547
Kernel: rbf C: 1 Gamma: 10 Accuracy: 0.8251526803890522
Kernel: rbf C: 100 Gamma: 0.1 Accuracy: 0.8453969690115358

*** The code generates many more entries, but in the interest of space, I only included about half of them in the write up.

Most accurate: **Kernel:** linear **C:** 1 **Accuracy:** 0.8548970821081203

Q3 - Sample Exam Questions

1)

- (True/False) Number of support vectors do not depend on the selected value of C. Please provide a one-sentence justification.

  Answer: False - C determines how many outliers will be taken into account when determining the support vectors, and therefore the margin is affected by C, so since we know that a narrower margin will have fewer support vectors and a wider margin will have more support vectors, so C does affect the number of support vectors.

- Select the correct option:

  Answer: Option 2: $n\_c=0 < n\_c=\infty$. When C is increased, the penalty of violating the margin constraints is higher, narrowing the margin and leading to a smaller amount of support vectors. Therefore, a lower value for C will lead to fewer support vectors than a higher value of C.

- Match the following list of kernels used for SVM classification with the following figures:

  Answer: (From left to right) 2: polynomial kernel, 1: linear kernel, 3: radial basis kernel

- What is the leave-one-out cross-validation error for the maximum margin based separation in the following figure ? (we are asking for a number) Please provide a (at least one-sentence) justification.

Answer: The LOOCV error would be zero, because no matter which one point you choose to leave out, the maximum margin-based separation would be the same, and since no change results from leaving one out, the error is zero.
2.

2)

A. Are the positive examples linearly separable from the negative examples in the original space?

Answer: No - you can't separate the positive points from the negative points with a straight line so they aren't linearly separable.

B. Consider the feature transformation $\varphi(x) = [1, x1, x2, x1x2]$, where x1 and x2 are, respectively, the first and second coordinates of a generic example x. The prediction function is $y(x) = w\ T * \varphi(x)$ in this feature space. Give the coefficients , w, of a maximum-margin decision surface separating the positive examples from the negative examples. (You should be able to do this by inspection, without any significant computation.)

Answer: W = (0, 0, 0, 1) transposed, since we want x1 = x2 = 0 for our lines, with a margin of 1, since x1x2 is positive for all positive points and negative for all negative points.

C. What kernel K(x, x') does this feature transformation φ correspond to?

Answer: It corresponds with $K(x, x') = \varphi(x)*\varphi(x') = 1 + x1*x1' + x2*x2' + x1*x1'*x2*x2'$.